International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:1 Volume No.10 January - February – 2026
DOI:10.46647/ijetms.2026.v10i01.001 ISSN: 2581-4621

HYBRID DEEP LEARNING MODEL OF SKIN LESION DIAGNOSIS BASED ON RESNET 50 AND CAPSULE NETWORK

L. Naga Srinivasu[1] [0000-0002-4944-4847], V. Chandra Sekhar[2], P. Karthik[2], Ch. Renu Sowmya Priya[2], K. Deepthi[2], Ventrapragada Teju[3]

[1]Assistant Professor, Department of ECE, Dhanekula Institute of Engineering & Technology, Vijayawada, Andhra Pradesh, India – 521139.

[2]UG Student, Department of ECE, Dhanekula Institute of Engineering & Technology, Vijayawada, Andhra Pradesh, India – 521139.

[3]Assistant Professor, Department of ECE, NRI Institute of Technology, Pothavarappadu (V), Vijayawada, Andhra Pradesh-521212, India

Abstract. A common and potentially fatal condition, skin cancer requires early and precise identification in order to be effectively treated. This paper introduces a hybrid deep learning model based Resnet 50 and Capsule Network for automated skin lesion classification using dermoscopic images. The model follows multiple stages, including preprocessing, segmentation, feature extraction, and classification. Preprocessing involves resizing, data augmentation, and normalization to enhance model performance. The active contour snake model is applied for segmentation to isolate lesion areas, improving feature extraction accuracy. Segmented images are then processed using ResNet50 for deep feature extraction, capturing key Lesion characteristics include convexity, circularity, irregularity, textural patterns, and color aspects. A Convolution Block Attention Module (CBAM) is integrated to refine these features, enhancing lesion pattern focus. The final classification module categorizes lesions into eight types, including melanoma and basal cell carcinoma. This framework enhances diagnostic accuracy, supporting dermatologists in early skin cancer detection and reducing subjectivity in manual diagnosis through advanced deep learning techniques.

Keywords: RESNET 50, Capsule Network, dermoscopic images, Convolution Neural Network.

1. Introduction

In particular, Deep Convolution Neural Networks (DCNNs) and Deep Neural Networks (DNNs) [1], are very effective machine learning models used for medical picture categorization and segmentation [2], [3]. DCNNs can learn complicated Representations and patterns derived from input images. This has led to greater utilization in identifying diseases like cancer. [4].Alzheimer's [5], cardiovascular disorders [6], and so forth. Although DCNNs are highly accurate, one major difficulty is their lack Of Interpretability [7]. These systems are commonly referred to as "black boxes" because of their difficulty in predicting outcomes. Formal medical practitioners may struggle to grasp the logic behind model outcomes, making it difficult to trust and depend on them. Interpretability is crucial in the medical industry, as machine learning conclusions might have catastrophic repercussions [8]. Understanding how a model reaches its conclusion can be aid in medical diagnosis.

Professionals recognize flaws, biases, and limits in model design or training [9]. This knowledge can help improve model accuracy and reliability and enable medical professionals to make educated judgments. Efforts are being made to tackle the challenge of interpretation in DCNNs [10].

The saliency map [11] identifies locations of interest by utilizing the activation values of various layers. By calculating the output's gradient in relation to the input image, these regions can be found. The magnitude of the gradient indicates the importance of each pixel in the picture for the

DOI:10.46647/ijetms.2026.v10i01.001 ISSN: 2581-4621

output prediction. By producing a spatial map of the areas that contribute most to a particular prediction, Class Activation Maps (CAM) were developed to determine the localized attention of the DCNNs [12]. It works by superimposing a global average pooling layer over the DCNN's last convolution layer. The class activation map was calculated using the weights that this layer had learned. The CAM is expanded upon by the Gradient Class Activation Mapping .This study suggests a brand-new, self-interpretable three-tier. It works by covering the last convolution layer of the DCNN with a global average pooling layer. The class activation map was calculated using the weights that this layer had learned. The CAM technique is further upon by Gradient Class Activation Mapping (Grad CAM), which computes the class activation map using gradient information [13].

The gradient from the last convolution block was used to weight feature maps, yielding a visual representation of the most relevant regions for prediction. Numerous improvements have been made to these techniques, such as using several layers to create the saliency map and guided back propagation to make them easier to understand[14].The use of other visualization techniques, such occlusion-based methods, which include masking parts of the input image in order to analyze the DCNN's output, has also been studied by researchers. There have been few attempts to explicitly integrate post-hoc interpretation techniques into model architectural design to enhance model interpretability, despite their usefulness as previously mentioned[15].

In order to increase the prediction accuracy of the core model, this study proposes a three-tier, self-interpretable deep neural network architecture for illness classification that leverages post-hoc interpretation models like Grad CAM or Saliency maps as attention modules. Global, attention, and local are the three branches of the proposed categorization paradigm. The first learning step, known as the global branch, draws high-level features from input data. For classification tasks, it records traits and patterns that are helpful. The learnt global branch's post-hoc attention maps are used by the attention branch to carry out unsupervised segmentation. Key elements in input data that influence categorization decisions are highlighted by attention maps. The attention branch is used by the local branch to hone information from the global branch. The highlighted portions of the attention branch are exploited. To obtain more specific details on the disease under classification. In order to properly categorize illnesses, the local branch makes an effort to collect crucial information. By leveraging global data and concentrating on confined regions of interest, the local branch increases classification accuracy. The suggested method can improve classification performance once the model reaches a plateau in training. The suggested architecture increases the model's accuracy and clarifies its prediction-making process for researchers and practitioners. In industries like as healthcare, where precision and openness are equally important, this degree of interpretability is especially beneficial.

2. Proposed Methodology

In this paper, hrbrid deep learning model was developed to detect the various skin diseases. The block diagram of the proposed model is shown in figure 1. The proposed model is implemented by using various modules.

2.1 Dataset Acquisition

This paper uses publicly available datasets, ISIC 2020 [16] and HAM10000 [17], which are widely used for skin lesion classification tasks.These datasets contain dermoscopic images of different skin lesions, including normal skin, benign, and malignant conditions.

2.2 Dermoscopic Image Input

Dermoscopic images are high-resolution skin images captured using a dermatoscope, a specialized optical device that provides detailed visualization of skin structures beneath the surface. These images are crucial in dermatology, especially for early detection of skin cancers such as melanoma, as they reveal patterns and features not visible to the naked eye. Dermoscopy enhances diagnostic

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:1 Volume No.10 January - February – 2026
DOI:10.46647/ijetms.2026.v10i01.001 ISSN: 2581-4621

accuracy by allowing dermatologists and AI models to analyze lesion characteristics, including asymmetry, border irregularities, color variations, and textural patterns. Capturing these images can be done through contact dermoscopy, where the device is placed directly on the skin with a gel to reduce reflections, or non-contact dermoscopy, which uses polarized light to enhance lesion visibility without direct skin contact.

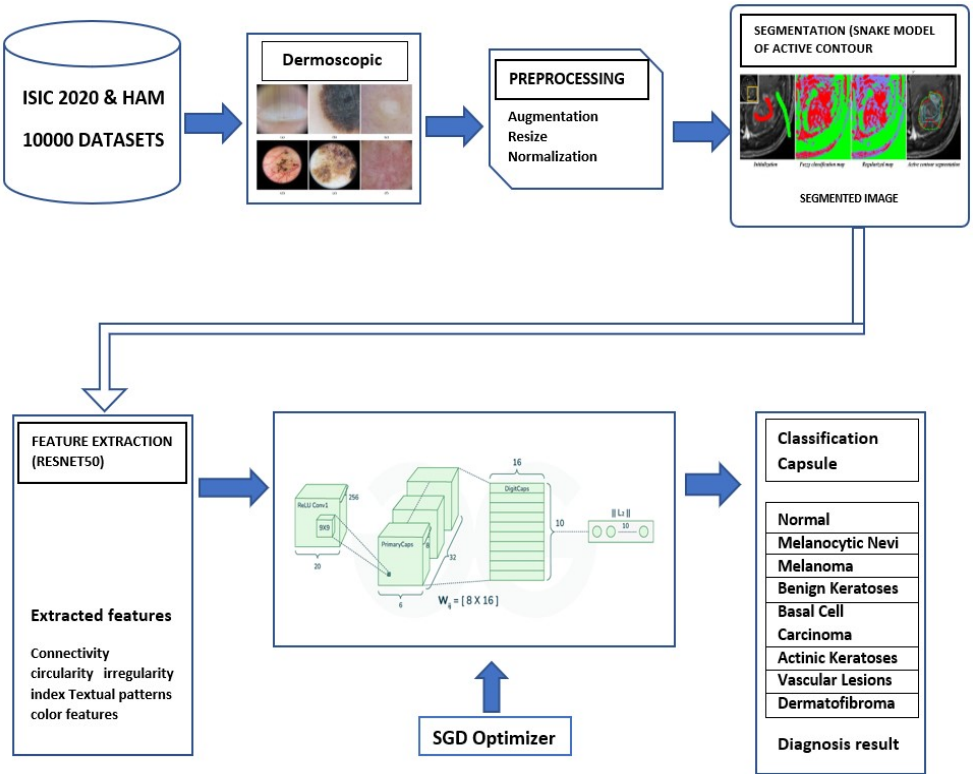

Fig. 1. Block Diagram of the Proposed Model

2.3 Pre-processing

Pre-processing [18] is a crucial step in dermoscopic image analysis, aimed at improving image quality, enhancing features, and standardizing data before further processing. Since raw dermoscopic images often contain noise, variations in lighting, artifacts (such as hair, air bubbles, and reflections), and differences in resolution, pre-processing techniques help refine the images for more accurate segmentation, feature extraction, and classification. The pre-processing pipeline typically includes resizing, data augmentation, normalization, and artifact removal, ensuring that the deep learning model can generalize well across different datasets and real-world conditions.

Resizing. It is essential to standardize input dimensions, as images in public datasets like ISIC 2020 and HAM10000 may have varying resolutions. Resizing ensures that all images are compatible with the neural network architecture while maintaining essential lesion details.

Data augmentation. Methods like color leveling, contrast correction, flipping, and rotation, Enhance training variety. samples, preventing overfitting and improving model robustness. This is particularly important because some skin lesion types are underrepresented in datasets, and augmentation helps balance class distribution.

Normalization. It is applied to adjust pixel intensity values to a consistent range, improving model convergence during training. This step often involves scaling pixel values between 0 and 1 or using z-score normalization to standardize brightness and contrast levels across images.

2.4. Segmentation

Segmentation[19] is a crucial step in dermoscopic image analysis, aimed at accurately isolating the skin lesion from the surrounding healthy skin One of the widely used segmentation techniques in

DOI:10.46647/ijetms.2026.v10i01.001 ISSN: 2581-4621

dermoscopic image processing is the Active Contour Model (Snake Model), which iteratively refines the boundary of the lesion by minimizing energy functions based on image gradients and region-based characteristics. This method effectively adapts to complex lesion shapes while avoiding over-segmentation caused by noise or artifacts. Other traditional segmentation methods include thresholding, which separates the lesion based on intensity values, and edge detection algorithms such as the Canny or Sobel operators, which identify lesion boundaries by detecting gradient changes. However, these methods may struggle in cases where the contrast between the lesion and the surrounding skin is low.

2.5. Feature Extraction (ResNet50 + CBAM)

Feature extraction is a critical step in skin lesion classification, where deep learning models extract meaningful patterns from dermoscopic images to differentiate between various skin conditions. In this study, ResNet50[20] (Residual Network with 50 layers) is utilized as the primary feature extractor due to its deep architecture and residual learning capabilities, which help in overcoming vanishing gradient issues and improving model performance. ResNet50 employs multiple convolutional layers with skip connections, preserving spatial hierarchies and enabling the network to collect both low-level and high-level lesion characteristics.Through this process, Convexity, circularity, irregularity index, textural patterns, and colour traits are among the most important qualities retrieved. These features help in analyzing the lesion's shape, structure, pigmentation, and surface variations, which are critical for distinguishing between benign and malignant cases.

A Convolutional Block Attention Module (CBAM)[21] is incorporated into the framework to improve feature representation even further.CBAM applies spatial and channel attention mechanisms, helping the model focus on the most relevant lesion regions while filtering out unnecessary background noise. The channel attention mechanism refines feature maps by emphasizing important channels, while the spatial attention mechanism ensures that key lesion regions are highlighted. By incorporating CBAM, the system improves its ability to accurately detect and classify skin lesions, leading to enhanced diagnostic reliability.

2.6. Capsule Network for Classification

Capsule Network [22] consists of multiple layers, beginning with a convolutional layer that extracts low-level features such as edges and textures. These features are then passed to the Primary Capsule Layer, where they are transformed into higher-dimensional vectors representing different attributes of an object. The most distinctive part of CapsNet is dynamic routing by agreement, which replaces max pooling and ensures that only the most relevant information is passed forward. Instead of simply activating a neuron based on the highest feature response, CapsNet dynamically routes information based on feature agreement, leading to a more accurate and robust classification. The final Digit Capsule Layer aggregates these transformed vectors to make predictions, capturing the hierarchical structure of an object rather than treating features independently**.**

When combined with deep feature extraction models such as ResNet50 and attention mechanisms like CBAM (Convolutional Block Attention Module), CapsNet further enhances classification accuracy. Its ability to retain spatial hierarchies, model complex transformations, and improve robustness against noise makes it a powerful tool for automated dermatological diagnosis, assisting dermatologists in making more reliable and early cancer detections.

2.7. Classification and Diagnosis

The final classification output includes different categories of skin lesions:

Normal. Refers to healthy skin without any abnormal growths, pigmentation issues, or signs of malignancy.

Melanocytic Nevi. These benign growths, which are made up of melanocytes, or cells that produce pigment, are also referred to as moles and can vary in size, shape, and color.

Melanoma. A highly aggressive and dangerous form of skin cancer that arises from melanocytes, often characterized by asymmetrical, irregularly bordered, and multicolored lesions.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:1 Volume No.10 January - February – 2026
DOI:10.46647/ijetms.2026.v10i01.001 ISSN: 2581-4621

Benign Keratosis. A non-cancerous skin growth that appears as rough, scaly, or wart-like patches, commonly associated with aging and sun exposure.
Basal Cell Carcinoma (BCC). Usually slow-growing and locally invasive, the most prevalent kind of skin cancer frequently manifests as a pearly or waxy lump on places exposed to the sun.
Actinic Keratoses. Long-term sun exposure can induce precancerous skin lesions, which manifest as rough, scaly areas and, if left untreated, can progress to squamous cell carcinoma.
Vascular Lesions. Abnormal blood vessel formations, such as hemangiomas and spider veins, which can be congenital or acquired and may appear red, blue, or purple on the skin.
Dermatofibroma. A benign skin tumor composed of fibrous tissue, often appearing as a small, firm, and slightly pigmented nodule that may be tender when pressed.
The diagnosis result is based on the classification output.

2.8. SGD Optimizer

The Stochastic Gradient Descent (SGD)[23] optimizer is a fundamental optimization algorithm used in deep learning to minimize the loss function and improve model performance.This variation of gradient descent iteratively modifies model parameters according to the gradient of the loss function for each parameter.Unlike traditional batch gradient descent, which computes gradients over the entire dataset, SGD updates parameters using a single data sample (or a small batch) at a time, making it computationally efficient and suitable for large-scale datasets. This incremental updating process helps models converge faster and prevents excessive memory usage.

3. Results and Discussions

Discuss the performance and outcomes of the suggested model in this area. The performance measures aid in estimating the suggested model's performance. The performance is estimated in this study using the Accuracy, Precision, Recall, and F1-Score criteria.
Accuracy. Measures the overall correctness of a model by calculating the proportion of correctly classified samples out of the total samples.
Precision. Demonstrates how well the model is in identifying a particular class by displaying the percentage of properly predicted positive cases out of all projected positive cases.
Recall. Recall, which is often referred to as sensitivity or true positive rate, quantifies the percentage of real positive cases that the model accurately detected, demonstrating how well it detects positive occurrences.
F1 Score. When working with unbalanced datasets, a single performance measure that balances precision and recall is very helpful.
The proposed model was compared with the various existing trained models such as Resnet 50, Adam Resnet 50,

Table 1. Proposed model compared with various existing trained model

Algorithm name	Accuracy	Precision	Recall	F1 Score
Existing ResNeXt50 Adam	63.914	63.998	65.018	55.587
VGG19 Adam	56.574	56.305	57.024	55.024
Existing ResNet50 Adam	43.119	28.783	43.931	34.759
Proposed Hybrid Model	98.165	98.256	98.128	98.174

Table 1 compares the accuracy, precision, recall, and F1 score of four deep learning models: ResNeXt50 with Adam optimizer, VGG19 with Adam optimizer, ResNet50 with Adam optimizer, and a Proposed Hybrid Model.The accuracy, precision, recall, and F1 score of the current ResNeXt50 Adam model are 63.998%, 65.018%, and 63.914%, respectively. The VGG19 Adam model performs slightly lower, attaining 56.574% accuracy, 56.305% precision, 57.024% recall, and an F1 score of 55.024%. The Existing ResNet50 Adam model shows the weakest performance, with only 43.119% accuracy, 28.783% precision, 43.931% recall, and an F1 score of 34.759%. In contrast, the Proposed Hybrid Model significantly outperforms all the existing models, achieving an

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:1 Volume No.10 January - February – 2026
DOI:10.46647/ijetms.2026.v10i01.001 ISSN: 2581-4621

impressive accuracy of 98.165%, precision of 98.256%, recall of 98.128%, and an F1 score of 98.174%. This substantial improvement suggests that the hybrid approach effectively enhances the model's classification ability, demonstrating superior generalization and robustness compared to traditional architectures. The figure 2-6 shows the simulation results of the proposed method. From the results, we conclude that the proposed method indentifies the skin lesion disease effectively.

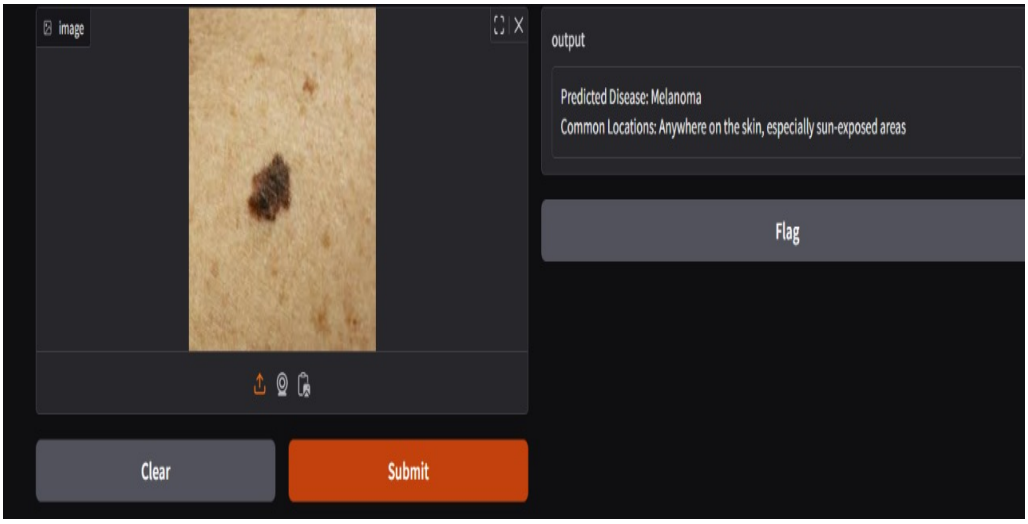

Fig. 2: MelanomaDisease Detected by Proposed Model

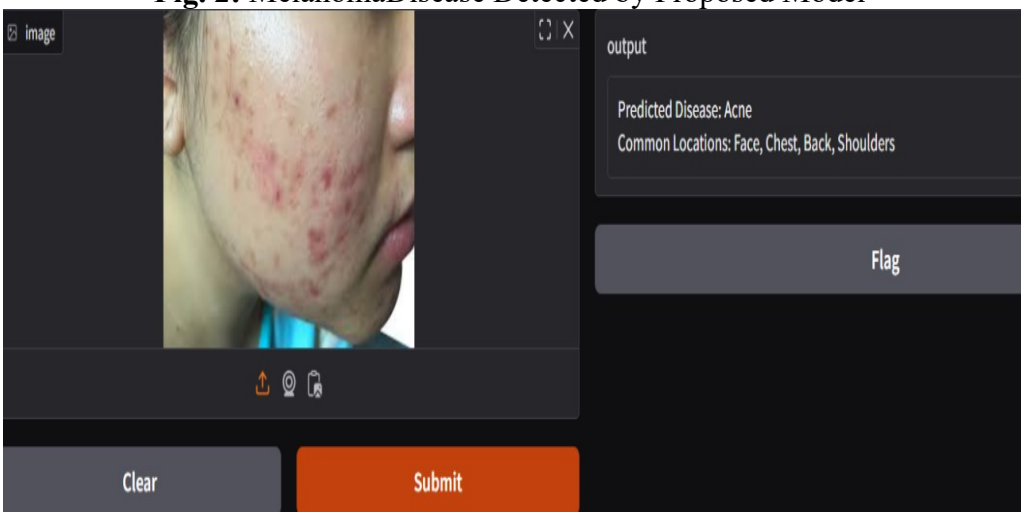

Fig. 3. Acne Disease Detected by Proposed Model

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:1 Volume No.10 January - February – 2026
DOI:10.46647/ijetms.2026.v10i01.001 ISSN: 2581-4621

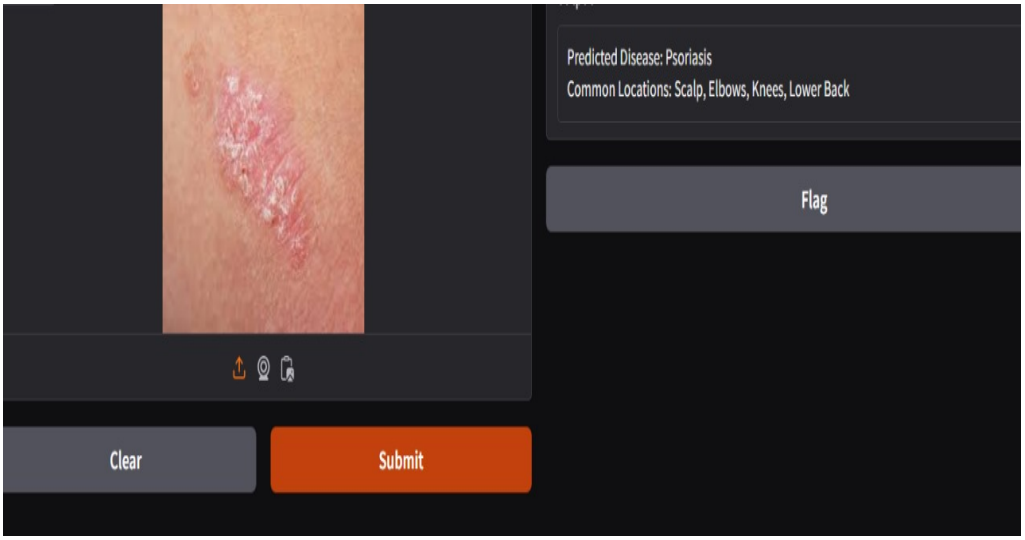

Fig. 4. Psoriasis Disease Detected by Proposed Model

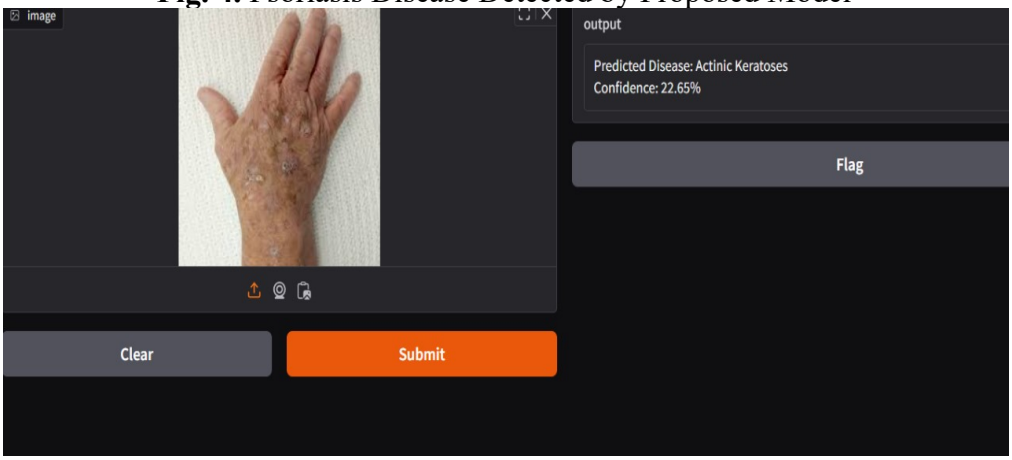

Fig. 5. Actinic Keratoses Disease Detected by Proposed Model

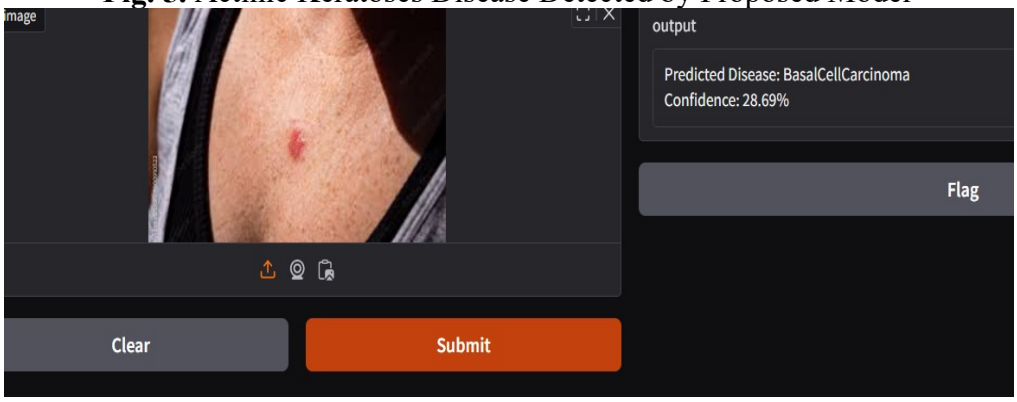

Fig. 6. Basal Cell Carcinoma Disease Detected by Proposed Model

4. Conclusion

The suggested deep learning-based architecture for classifying skin lesions provides an automated and extremely effective method for detecting skin cancer early on.By leveraging dermoscopic images from the ISIC 2020 and HAM10000 datasets, the system ensures a rich and diverse dataset for training. The preprocessing stage enhances the quality of images through resizing, augmentation, and normalization, improving model generalization.The active contour snake model-based segmentation procedure isolates the lesion area, improving the precision of feature extraction.

DOI:10.46647/ijetms.2026.v10i01.001 ISSN: 2581-4621

Important lesion characteristics including convexity, circularity, irregularity index, textural patterns, and color data are recovered by the feature extraction process driven by ResNet50.By concentrating on important patterns, the convolutional block attention module (CBAM) further enhances these retrieved characteristics.TheCapsule Network, integrated with dynamic routing, replaces traditional convolutional layers, preserving spatial relationships between features and improving classification robustness. Finally, the model is optimized using the Stochastic Gradient Descent (SGD) optimizer, ensuring efficient learning and fast convergence.

A variety of skin lesions, such as melanocytic nevi, melanoma, benign keratosis, basal cell carcinoma, actinic keratoses, vascular lesions, and dermatofibroma, are correctly identified by the last categorization module. In the end, this paradigm improves patient outcomes by greatly increasing diagnostic accuracy and offering dermatologists dependable assistance with early diagnosis and treatment planning.

References

1. LeCun, Y., Bengio, Y., & Hinton, G.: Deep learning. Nature, 521(7553), 436-444 (2015).
2. Litjens, G., Kooi, T.,Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., and van der Laak, J. A. W. M.: A survey on deep learning in medical image analysis. Medical Image Analysis 5 (11), (2017).
3. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118, (2017).
4. Suk, H.-I., Lee, S.-W., Shen, D., and Alzheimer's Disease Neuroimaging Initiative.: Deep ensemble learning for Alzheimer disease classification. IEEE Access 6(3), 97-102, (2017).
5. Kwon, D., Kim, B., Kim, C., and Park, K.: Deep learning approaches for cardiovascular disease detection: A review. IEEE Access, 7, 84877-84888 (2019).
6. Doshi-Velez, F., and Kim, B.: Towards A Rigorous Science of Interpretable Machine Learning. *arXiv preprint arXiv:1702.08608,* (2017).
7. Lipton, Z. C.: The Mythos of Model Interpretability. Nature 2(1), 67-73, (2016).
8. Ribeiro, M. T., Singh, S., & Guestrin, C.: Why Should I Trust You? Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144, (2016).
9. Montavon, G., Samek, W., & Müller, K. R.: Methods for interpreting and understanding deep neural networks.Digital Signal Processing,3(5), 12-20, (2018).
10. Tjoa, E., and Guan, C.: A survey on explainable artificial intelligence (XAI): Towards medical AI transparency. Nature Machine Intelligence, 2, 56-66, (2020).
11. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. Proceedings of the IEEE International Conference on Computer Vision, 618-626, (2017).
12. Simonyan, K., Vedaldi, A., and Zisserman, A.: Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. arXiv preprint arXiv:1312.6034, (2013).
13. Sundararajan, M., Taly, A., and Yan, Q.: Axiomatic attribution for deep networks. Proceedings of the 34th International Conference on Machine Learning, 3319-3328, (2017).
14. Zhang, Q., Yang, Y., Ma, H., and Wu, Y.: Interpretable Deep Learning in Medical Image Analysis: Concepts, Methods, and Applications. Artificial Intelligence in Medicine, 122, 102157, (2021).
15. Chefer, H., Gur, S., and Wolf, L.: Transformer interpretability beyond attention visualization. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 782-791, (2021).
16. 16. ISIC 2020 Dataset https://challenge2020.isic-archive.com/
17. HAM10000 Dataset

DOI:10.46647/ijetms.2026.v10i01.001 ISSN: 2581-4621

https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T
18. Kass, M., Witkin, A., and Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision 2(3) 21-33, (1988).
19. Naga Srinivasu, L., Sumanth, P., and Srinivsarao, K.: Real time implementation of LWT based novel IR and VI fusion algorithm using raspberry pi. Telecommunictions and radio engineering,83(4), (2024).
20. He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778 (2016).
21. Woo, S., Park, J., Lee, J. Y., and Kweon, I. S.: CBAM: Convolutional Block Attention Module. Proceedings of the European Conference on Computer Vision (ECCV), 3-19, (2018).
22. Sabour, S., Frosst, N., and Hinton, G. E.: Dynamic Routing Between Capsules. Advances in Neural Information Processing Systems (NeurIPS), 30.arXiv:1710.09829, (2017).
23. Bottou, L.: Large-scale machine learning with stochastic gradient descent. Proceedings of the 19th International Conference on Computational Statistics (COMPSTAT), 177-186, (2010).